Design of one-step and multistep adaptive of algorithms for the tracking of time varying systems

Albert Benveniste

CENTRE DE RENNES
IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP105
78153 Le Chesnay Cedex
France
Tél: (3) 954 90 20

Rapports de Recherche

N° 340

DESIGN OF ONE-STEP AND MULTISTEP ADAPTIVE ALGORITHMS FOR THE TRACKING OF TIME VARYING SYSTEMS

Albert BENVENISTE

Septembre 1984

DESIGN OF ONE-STEP AND MULTISTEP ADAPTIVE ALGORITHMS FOR THE TRACKING OF TIME VARYING SYSTEMS

Albert BENVENISTE
IRISA/INRIA
Campus de Beaulieu
F-35042 RENNES CEDEX
FRANCE

Publication Interne n°236
Septembre 1984
40 pages

ABSTRACT: The design of adaptive algorithms for the purpose of the tracking of slowly time varying systems is investigated. A criterion for measuring the tracking capability of an algorithm in this situation was introduced in an earlier work; the domain of validity of this criterion is shown to be much wider than expected before. On the other hand, multistep algorithms, introduced in the Soviet litterature, are generalized and systematically studied; they are shown to provide significant improvements over the classical (one-step) methods for the purpose of tracking. Finally, a complete design methodology for adaptive algorithms used on time varying systems is given.

CONCEPTION D'ALGORITHMES ADAPTATIFS MONO- ET MULTIPAS POUR LA POURSUITE DE SYSTEMES NON STATIONNAIRES

RESUME: Dans ce papier, on examine les problèmes posés par l'utilisation d'algorithmes adaptatifs pour la poursuite de systèmes lentement variables dans le temps. On introduit un critère de qualité permettant d'évaluer à priori la capacité de poursuite d'un algorithme. D'autre part, l'on s'intéresse aux algorithmes multipas introduits dans les travaux de l'école Russe; ces algorithmes multipas présentent des intérêts pour la poursuite de systèmes non stationnaires. Enfin, on donne une méthodologie complète pour la conception d'algorithmes adaptatifs dans ce contexte.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(L.A. 227)
UNIVERSITÉ DE RENNES 1 I.N.S.A. DE RENNES

INSTITUT NATIONAL DE RECHERCHE
EN INFORMATIQUE ET EN AUTOMATIQUE
(LABORATOIRE DE RENNES)

PART I: INTRODUCTION, EXAMPLES, AND THEORETICAL BACKGROUND.

Among several other reasons, the success of adaptive algorithms in signal processing and automatic control is due to their ability to track slowly time varying systems with time invariant models. But, surprisingly, little has been done in order to obtain a convenient theoretical framework for analysing this problem. In order to define more precisely what we have in mind with the "tracking of slowly time varying systems", let us recall the problems that are encountered in the design and analysis of adaptive algorithms.

The first problem is the problem of **convergence** (or **consistency** in the framework of statistics): in this case, the true system is assumed to be time invariant. This is a wellknown problem we will not talk about.

The second problem is the **transient behaviour**: how fast the adjusted parameter converges in the vicinity of the desired value after an initialization, or after some abrupt and "large" change in the true system? To my knowledge, only rules of thumb are now available to handle this problem.

The third problem is the **convergence rate**: the true system is assumed to be time invariant, and one is interested in the speed of convergence of the algorithm to the desired value of the parameter after a long period of time. Extensive studies have been done in this direction, in the litterature of statistics as well as in the litterature of identification. Again, this is not the purpose of this paper.

Finally, the **tracking** problem we want to investigate in this paper is the following: assume that, after some transient period, the adjustable parameter has approached to the vicinity of its desired "true" value. We want now to investigate what happens if the true system is varying "slowly" with time. It has been recognized for a long time that a good way of modifying the classical adaptive algorithms (i.e. with decreasing gains) is, either to introduce a forgetting factor, or to use constant gain algorithms (see Ljung & Soderström (1983)). Another way is to set the identification problem in such a way that the tracking is obtained through the use of an extended Kalman filter (see Bohlin (1977) for a first attempt in this direction).

A first family of papers has been involved in getting upper bounds of the short time mean square error (by error, we have in mind the deviation of the estimated from the actual value of the parameter), by only assuming that the speed of variation of the true system is bounded by some deterministic constant. The work of Farden & Sayood (1980) belongs to this family, where the "Widrow" LMS algorithm is investigated in the case of independent regression vectors. But the most interesting contribution has been given by Macchi & Eweda (1981) for the same algorithm, but for m-dependent regression vectors and with more illuminating results. In fact, I should say that this paper is the only rigorous treatment of this problem which avoids the use of any asymptotic argument, but this method applies to this simple algorithm only.

A second family of papers uses asymptotic arguments for the prupose of the analysis, thus resulting in questions like "what is the domain of validity of the approximation?", but with more general methods, and results leading to effective design methods for adaptiva algorithms. To this family belongs the pioneering work of Widrow & al.(1976) where the LMS algorithm is analysed for independent regression vectors. Let us also mention the work of Benveniste & Ruget (1982) for general adaptive algorithms with constant gains. Related to this approach is also the work of Kushner & Huang (1981) where a diffusion model is given for describing the behaviour of a general adaptive algorithm in the presence of time varying systems, and of Kushner & Huang (1982), where the digital phaselock loop is analysed. Finally, a recent paper of Widrow & Walach (1984) presents results similar to those of Widrow & al (1976). The present paper also belongs to this class; compared to our previous work, we shall introduce our asymptotic analysis in a more satisfactory way, thus resulting in a large possibility of generalization.

As we shall see in this paper an interesting generalization will be the analysis of **multistep adaptive algorithms** in the case of the tracking of time varying systems. Multistep methods for the finite difference approximation of differential equations are wellknown in the field of numerical analysis, see Henrici (1963) for a classical treatment. Howewer, multistep algorithms have been also introduced in the field of stochastic approximations of the Robbins-Monro type by several authors of the Russian school (Shil'man & Yastrebov (1976, 1978), and the remarkable paper of Korostelev (1981) where multistep algorithms are analysed from the viewpoint of the large deviation theory), and by Ruszczynskyi & Sysky (1983). Shil'man & Yastrebov have investigated in a heuristic way and by simulations the transient behaviour of multistep algorithms, showing that interesting improvements can be obtained in some cases. But a long discussion occurred in the Soviet litterature about the convergence rate of those procedures: some of them are shown to be less efficient, and other as efficient as the classical one-step methods. One of the pruposes of the present paper is to clarify this discussion.

The purpose of the present paper is **to give quantitative design methods for both one-step and multistep adaptive algorithms in the case of the tracking of slowly time varying systems.** Although the approach is fairly general, the results are more relevant to adaptive signal processing thant to adaptive control. The questions to which we shall give an answer are the following:

(i) Is it possible to measure a priori (i.e; without any prior knowledge of the possible disturbance acting on the true system) **the tracking capability of a given adaptive algorithm?**

(ii) How to design in an optimal way the algorithm when the characteristics of the disturbance are approximately known? (Optimal design means here the choice of a one- or multistep form of the algorithm, and of the corresponding gains).

The paper is organized as follows. First of all, two basic examples will be introduced for the purpose of the illustration of our approach; then the necessary theoretical background will be presented.

The second part of the paper will be devoted to the analysis of one-step algorithms, i.e. classical ones. A new asymptotic analysis will be presented, and answers will be given to the two questions mentioned above. The tradeoff tracking/accuracy will be analysed for both zero- and non zero drift models of the time variations of the true system. A link will be established between the measure of the tracking capability, and the classical Fisher information matrix.

The third part of the paper will be devoted to multistep methods. Multistep algorithms will be shown to improve the tracking capability, and the problem of the optimal design of a multistep adaptive algorithm will be solved for non zer drift models of the time variations of the true system. A connection will be established with the Kalman filter approach and the "smoothness priors" of Gersh & Kitagawa (1984).

I TWO TYPICAL EXAMPLES.

I.1 THE DIGITAL PHASE LOCKED LOOP.

We refer the interested reader to Benveniste, Vandamme & Joindot (1980) for a more detailed presentation. Let us consider the case of the 4-PSK (Phase Shift Keying) transmission scheme in data communication; we shall describe this system in the baseband equivalent form (Falconer (1976)). An i.i.d. complex signal $(a_t)$of the following form

$$a_t = e^{i\phi_t}, \quad \phi_t \in \{ \pi/4 + k2\pi/4, \quad k=0,1,2,3 \} \tag{1.1}$$

is sent through a complex channel; the receiver observes the complex signal

$$y_t = (\sum_{k\in Z} s_k a_{t-k}) e^{i\phi*} + v_t \tag{1.2}$$

where (s_k) is the channel (including the baseband equivalent effect of the transmission and reception filters, together with the effect of the proper channel, the noise v being also complex); ϕ_* is the phase shift due to the channel. The model (1.2) is redundant unless we assume the normalizing condition $s_0 > 0$. Assuming that the distorsion of the channel is small, i.e.

$$\sum_{k\neq 0} |s_k| << s_0 \tag{1.3}$$

the main degradation is then due to the unknown phase shift ϕ_*. The purpose of the DPLL is then **to estimate this phase shift, to rotate the received signal by the opposite of the corresponding estimate ϕ , and to apply a simple decision rule to $y_t . e^{-i\phi_t}$ for recovering an estimate $\hat{a}$ of the message.** Typical algorithms for estimating ϕ are the following

$$\phi_{t+1} = \phi_t - \gamma\varepsilon_t(\phi_t), \quad \gamma>0 \tag{1.4}$$

where

$$\varepsilon_t(\phi) = \mathrm{Im}(y^4 e^{-i4\phi}) \qquad \text{(Costas Loop)}$$

$$\varepsilon_t(\phi) = -\mathrm{Im}(y_t e^{-i\phi} \hat{a}_t^*(\phi)) \qquad \text{(Decision Feedback Loop)} \tag{1.5}$$

where $\hat{a}_t(\phi)$ is given by

$$\hat{a}_t(\phi) = \mathrm{sign}(\mathrm{Re}\, y_t e^{-i\phi}) + i\, \mathrm{sign}(\mathrm{Im}\, y_t e^{-i\phi}) \tag{1.6}$$

(the superscript * denotes the complex conjugate).

I.2 THE LEAST SQUARES ALGORITHM.

Let (y_t) be a signal of the form

$$y_t = \phi_t^T.\theta_* + v_t \tag{1.7}$$

where (ϕ_t) is an n-dimensional regression vector, and (v_t) is a zero mean process such that v_t is independent of ϕ_t. The Recursive Least Squares (RLS) algorithm with constant matrix gain is given by

$$\begin{aligned} \theta_{t+1} &= \theta_t + A\,\phi_t\, e_t(\theta_t) \\ e_t(\theta_t) &= y_t - \phi_t^T.\theta_t \end{aligned} \tag{1.8}$$

where A is a constant matrix gain to be chosen.

II THEORETICAL PEREQUISITES.

The general form of adaptive algorithms we can use for the two examples introduced above is the following

$$\theta_{t+1} = \theta_t + A\,V_t(\theta_t) \quad , \quad \theta \in \mathbf{R}^n \tag{1.9}$$

where A is a constant matrix gain, and, for θ fixed, $(V_t(\theta))_{t\in\mathbf{Z}}$ is an ergodic stationary random vector field; less restrictive forms are also used in the litterature, see for instance Lung & Soderström (1983), and Métivier & Priouret (1984), where a more general conditional Markov model is used. For the sake of simplicity, we shall restrict ourselves to the case of stationary random vector fields, but the method we use extends without any modification to the more general models we mentioned above. The classical tools for analysing the algorithm (1.9) are now given.

For the prupose of the analysis of the small gain case, let us modify (1.9) by introducing a small parameter :

$$\theta_{t+1} = \theta_t + \gamma\,A\,V_t(\theta_t) \tag{1.10}$$

II.1 THE ORDINARY DIFFERENTIAL EQUATION (ODE).

The ODE associated to (1.10) is defined by

$$\dot{\theta} = A\,V(\theta) \quad , \quad V(\theta) = \mathbf{E}(V_t(\theta)) \tag{1.11}$$

and we shall denote by

$$\bar{\theta}(\tau)\ ,\ \tau \geq 0 \tag{1.12}$$

the solution of the ODE such that $\bar{\theta}(0)=\theta_0$. Then we have the following result:

Theorem 1: let T be finite or infinite. Then, for every positive ε , we have

$$\lim_{\gamma \to 0} P\{ \sup_{0 \leq t\gamma < T} ||\theta_t - \bar{\theta}(t\gamma)|| > \varepsilon \} = 0. \tag{1.13}$$

Without any further assumption on the ODE, we must assume T finite in this theorem (Benveniste & al. (1980)), whereas we can assume T infinite if convenient stability conditions are satisfied by the ODE (Derevitskii & Fradkov (1974)). Typical assumptions for the theorem 1 to be valid are the following (precise statement can be found in the above mentioned references, and in Métivier & Priouret (1984)):

Assumptions for the theorem 1:

for θ fixed, $(V_t(\theta))$ is ergodic (1.14)

$\theta \rightarrow V_t(\theta)$ **can be discontinuous, but $\bar{\theta} \rightarrow V(\bar{\theta})$ is smooth** (1.15)

Related results for the more classical case of decreasing gain algorithms can be found for example in Ljung & Soderström (1983), and in Métivier & Priouret (1984).

II.2 THE INVARIANCE PRINCIPLE.

Fix again T finite; we shall now investigate in a more precise way the error

$$\theta_t - \theta(\tau)$$

(for $\tau = t\gamma$) as γ tends to zero. Let us denote by

$$(\tilde{\theta}^{\gamma}_{\tau})_{0 \leq \tau \leq T}$$

the continuous time stochastic process with piecewise linear trajectories such that

$$\tilde{\theta}^{\gamma}_{\tau} = \gamma^{-1/2} (\theta_t - \bar{\theta}(\tau)) \quad \text{for} \quad \tau = t\gamma \tag{1.16}$$

Then, we have the following invariance principle (Kushner & Huang (1981)):

Theorem 2: When γ tends to zero, the process $(\tilde{\theta}_\tau^\gamma)$ converges weakly to the Gaussian process $(\tilde{\theta}_\tau)$, which is the solution of the linear stochastic differential equation

$$d\tilde{\theta}_\tau = A\, V_\theta(\bar{\theta}(\tau))\, \tilde{\theta}_\tau\, d\tau + A\, R^{1/2}(\bar{\theta}(\tau))\, dW_\tau\,, \quad \tilde{\theta}_0 = 0 \tag{1.17}$$

where (W_τ) is a standard n-dimensional Brownian motion, and

$$\begin{aligned} V_\theta(\bar{\theta}) &= \frac{d}{d\theta} V(\bar{\theta}) \\ R(\bar{\theta}) &= \sum_{n\in Z} \operatorname{cov}(V_n(\bar{\theta}), V_0(\bar{\theta})) \end{aligned} \tag{1.18}$$

the serie being assumed to converge.

For this theorem to be valid, (1.14) has to be reinforced by requiring some mixing conditions on the random vector field $(V_t(\theta))$.

PART 2: MONOSTEP ADAPTIVE ALGORITHMS.

I TOWARDS A MODEL FOR ANALYSING THE TRACKING PROBLEM.

I.1 BACK TO THE EXAMPLES.

Let us first analyse the DPLL. We shall now take into account that the true phase shift of the channel is time varying, a situation which is most often encountered. Then it is convenient to rewrite (1.2) as follows

$$y_t(\phi_*(t)) = \Big(\sum_{k\in\mathbf{Z}} s_k\, a_{t-k} \Big)\, e^{i\phi_*(t)} + v_t \tag{2.1}$$

which turns out to modify the expression of the DPLL algorithms as follows

$$\phi_{t+1} = \phi_t - \gamma\, \varepsilon_t(\phi_t,\phi_*(t)) \tag{2.2}$$

where

$$\varepsilon_t(\phi,\phi_*) = \mathrm{Im}\big(y_t^4\, e^{-i4(\phi-\phi_*)} \big) \qquad \textbf{(Costas Loop)}$$

(2.3)

$$\varepsilon_t(\phi,\phi_*) = -\,\mathrm{Im}\big(y_t\, e^{-i(\phi-\phi_*)}\, \hat{a}_t^*(\phi-\phi_*) \big) \qquad \textbf{(Decision Feedback Loop)}$$

where y_t denotes $y_t(0)$ as defined by (2.1).

The same modification holds for the RLS algorithm when θ_* is time varying: (1.7) and (1.8) have to be replaced respectively by

$$y_t(\theta_*(t)) = \phi_t^T(\theta_*(t)).\theta_*(t) + v_t \tag{2.4}$$

and

$$\theta_{t+1} = \theta_t + A\,\phi_t(\theta_*(t)).e_t(\theta_t,\theta_*(t))$$

(2.5)

$$e_t(\theta,\theta_*) = y_t(\theta_*) - \phi_t^T(\theta_*).\theta$$

I.2 THE MODEL.

The analysis of the examples introduces in a natural way the modification of (1.9) which is required for the analysis on time varying systems:

$$\theta_{t+1} = \theta_t + A\, V_t(\theta_t, S_t) \tag{2.6}$$

where S denotes the parameter vector corresponding to the true system, and, for θ and S fixed, the random vector field $(V_t(\theta,S))_t$ is stationary.

For describing the time variations of the true system S, we shall use the fairly general model

$$S_{t+1} = S_t + U_t(S_t) \tag{2.7}$$

where, as usually, for S fixed, $(U_t(S))_{t\in Z}$ is a stationary random vector field. Since, as we have mentioned before, we are interested in the analysis of (2.6) when the true system moves slowly, we shall again introduce a small parameter γ in the joint recurrent equations (2.6) and (2.7), thus obtaining finally the model

$$\begin{aligned} \theta_{t+1} &= \theta_t + \gamma A\, V_t(\theta_t, S_t) \\ S_{t+1} &= S_t + \gamma\, U_t(S_t) \end{aligned} \tag{2.8}$$

and we shall study the behaviour of (2.8) when γ tends to zero.

Here, the adjustable matrix gain chosen by the designer is the matrix A, whereas the parameter γ is nothing but a tool for the analysis. Let us summarize the assumptions we need on (2.8):

Assumptions 1: For θ and S fixed, $(U_t(S))$ and $(V_t(\theta,S))$ are stationary random vector fields with means

$$U(S) = \mathbf{E}(U_t(S)), \quad V(\theta,S) = \mathbf{E}(V_t(\theta,S)). \tag{2.9}$$

We shall assume that the identification is perfect when $V(\theta,S)=0$ i.e.

$$V(\theta,S) = 0 \quad \text{iff} \quad \theta = S. \tag{2.10}$$

Finally, we shall assume the following condition on the first and second partial derivatives of V:

$$V_\theta(S,S) = -V_S(S,S). \tag{2.11}$$

(2.11) means that the two partial derivatives of V are opposite on the diagonal. This is in fact a very natural assumption. (2.11) is satisfied, for example, when $V(\theta,S)$ is of the form $V(\theta-S)$, which is for example the case for the DPLL in view of (2.3); but it is also satisfied by the RLS algorithm, as it can be easily verified, using the orthogonality condition between the noise and the regression vector on the true system.

Condition (2.10) could be weakened for allowing the case of overparametrization of the model with respect to the true system (in which case $V(\theta,S)=0$ has infinitely many solutions), but we shall restrict ourselves to (2.10) for the sake of simplicity.

I.3 THE BASIC PROBLEMS.

We shall now investigate the problems we mentioned in the introduction, namely:

Problem 1: how to choose in (2.8) the matrix gain A in an optimal way when the model of the time variations of the true system S is approximately known?

Problem 2: how to compare different random vector fields $V_t^1(\theta,S)$ and $V_t^2(\theta,S)$ used for the tracking of the same system, without any prior knowledge of the possible time variations of this true system?

Note that the later question occurs naturally when the purpose of the analysis is the selection of one of the two DPLL introduced above, according to the criterion of the best tracking capability.

A natural criterion which reflects the compromise between tracking and accuracy is the following: **try to keep the mean square deviation**

$$E||\theta_t - S_t||^2 \tag{2.12}$$

as small as possible when t is large (recall that we do not analyse here the transient behaviour of the algorithm). This criterion has already been used by most of the authors (Macchi & Eweda (1984), Benveniste & Ruget (1982), Farden & Sayood (1980)), whereas the criterion used by Widrow & al.(1976) will be shown to be very close to the present one.

I.4 EXPRESSING THE OBJECTIVE FUNCTION (2.12) AS THE SUM OF A BIAS PLUS A VARIANCE.

For $t\gamma=\tau$, let us rewrite θ_t-S_t in the following way

$$\theta_t - S_t = (\bar{\theta}(\tau) - \bar{S}(\tau)) + \gamma^{-1/2}(\tilde{\theta}^{\gamma}_{\tau} - \tilde{S}^{\gamma}_{\tau}) \tag{2.13}$$

where $(\bar{\theta},\bar{S})$ and $(\tilde{\theta}^{\gamma}_{\tau},\tilde{S}^{\gamma}_{\tau})$ are defined respectively according to (1.12) and (1.16) from the stochastic recurrent equations (2.8). In view of the theorems 1 and 2, for small, we have the approximation

$$\begin{aligned} \mathrm{Tr}\, E(\,(\theta_t-S_t)(\theta_t-S_t)^T) &\simeq \mathrm{Tr}(\,(\bar{\theta}(\tau)-\bar{S}(\tau))(\bar{\theta}(\tau)-\bar{S}(\tau))^T) \\ &\quad + \gamma\, \mathrm{Tr}(\,\mathrm{cov}(\tilde{\theta}_{\tau}-\tilde{S}_{\tau})\,) \\ &= \mathrm{Tr}(\,\mathrm{bias} + \gamma\ \mathrm{variance}\,). \end{aligned} \tag{2.14}$$

II MAIN THEOREMS.

Throughout this chapter, the following assumption will be in force:

Assumption 2: for every $\bar{S}$ fixed, the ODE

$$\dot{\theta} = A\, V(\bar{\theta},\bar{S}) \tag{2.15}$$

is asymptotically stable, with $\bar{\theta}=\bar{S}$ as unique equilibrium

The following quantities will be of interest:

$$\begin{aligned} Q(\bar{S}) &= \sum_{n\in Z} \mathrm{cov}(U_n(\bar{S}),U_0(\bar{S})) \\ R(\bar{S}) &= \sum_{n\in Z} \mathrm{cov}(V_n(\bar{S},\bar{S}),V_0(\bar{S},\bar{S})) \end{aligned} \tag{2.16}$$

(cf. theorem 2). When there is no possible ambiguity, we shall drop the dependence on the variable $\bar{S}$ in these quantities.

II.1 THE ZERO DRIFT CASE

It corresponds to the following assumption: the true system evolves as a zero mean random process, i.e. the ODE

$$\dot{\bar{S}} = U(\bar{S}),\ \bar{S}(0) = \bar{S}_\circ \tag{2.17}$$

has $\bar{S}=\bar{S}_\circ$ as solution. In this case, (2.15) and (2.16) give that **the bias is zero** (cf. (2.14)); in the sequel, when using the theorem 2, we shall drop the dependence on $(\bar{\theta},\bar{S})$, since these are constant in the zero drift case. We thus have only to investigate the variance. According to the theorem 2, the joint process $(\tilde{\theta},\tilde{S})$ is solution of the following stochastic differential equation

$$\begin{bmatrix} dS_\tau \\ d\theta_\tau \end{bmatrix} = \begin{bmatrix} U_S & 0 \\ AV_S & AV_\theta \end{bmatrix} \begin{bmatrix} S_\tau \\ \theta_\tau \end{bmatrix} d\tau + \begin{bmatrix} Q^{1/2} & 0 \\ 0 & AR^{1/2} \end{bmatrix} dW_\tau \tag{2.18}$$

where W_τ is a standard brownian motion, and U_S, V_S, and V_θ denote partial derivatives. We can here distinguish two subcases:

case (i): U_S is asymptotically stable.

This is for example the case if the true system moves according to a stable second order process. Then, using (2.11), we get

$$\begin{bmatrix} dS_\tau \\ \\ d\theta_\tau \end{bmatrix} = \begin{bmatrix} U_S & 0 \\ \\ -AV_\theta & AV_\theta \end{bmatrix} \begin{bmatrix} S_\tau \\ \\ \theta_\tau \end{bmatrix} d\tau + \begin{bmatrix} Q^{1/2} & 0 \\ \\ 0 & AR^{1/2} \end{bmatrix} dW_\tau \tag{2.19}$$

Finally, in this case (i), under the assumptions 1 and 2, the joint process $(\tilde{S}, \tilde{\theta}-\tilde{S})$ is the solution of the following linear stochastic differential equation

$$\begin{bmatrix} dS_\tau \\ \\ d\theta_\tau \end{bmatrix} = \begin{bmatrix} U_S & 0 \\ \\ -U_S & AV_\theta \end{bmatrix} \begin{bmatrix} S_\tau \\ \\ \theta_\tau \end{bmatrix} d\tau + \begin{bmatrix} Q^{1/2} & 0 \\ \\ -Q^{1/2} & AR^{1/2} \end{bmatrix} dW_\tau \tag{2.20}$$

Hence, setting

$$\underline{P} = \operatorname{cov}(\tilde{S}, \tilde{\theta}-\tilde{S}) \tag{2.21}$$

, $\underline{P}$ is the solution of the following Lyapunov equation

$$\underline{F}\,\underline{P} + \underline{P}\,\underline{F}^T + \underline{Q} = 0$$

$$\underline{F} = \begin{bmatrix} U_S & 0 \\ -U_S & AV_\theta \end{bmatrix} \qquad \underline{Q} = \begin{bmatrix} Q & -Q \\ -Q & Q + ARA^T \end{bmatrix} \tag{2.22}$$

where Q and R are defined in (2.16).

case (ii): $U_S = 0$.

This corresponds to the case where the true systems moves according to a random walk. In this case, (2.20) degenerates in the following equation

$$d\theta_\tau = AV_\theta\, \theta_\tau\, d\tau + (\, -Q^{1/2} \quad AR^{1/2})\, dW_\tau \tag{2.23}$$

which gives that

$$P = \operatorname{cov}(\tilde{\theta}-\tilde{S}) \tag{2.24}$$

is the solution of the Lyapunov equation

$$AV_\theta P + P(AV_\theta)^T + ARA^T + Q = 0. \tag{2.25}$$

II.2 THE NON ZERO DRIFT CASE

Here, the situation is more involved, since we shall have to take into account both bias and variance. We shall here assume that

$$U(\bar{S}) \neq 0 \text{ along the trajectory of (2.17).} \tag{2.26}$$

The following result is then proved in the appendix A: if (2.26) is in force, it is convenient to choose the matrix gain A so that the time variations of θ be much faster than the time variations of S. This is achieved in the following way: select A of the following form

$$A = \gamma^{-\alpha} A_o \tag{2.27}$$

where the matrix A_o is fixed, and α has to be chosen. Then, it is proved in the appendix A that **the optimum α is**

$$\alpha = 1/3 \tag{2.28}$$

Then it is proved that, **for γ small, and** $t\gamma=\tau$,

$$E||\theta_t - S_t||^2 \simeq \gamma^{2/3} \operatorname{Tr}((A_o V_\theta)^{-1} UU^T (A_o V_\theta)^{-T} + \mathbf{P_o}), \quad U \equiv U(\bar{S}(\tau)),$$

and $\mathbf{P_o}$ is the solution of (2.29)

$$A_o V_\theta \mathbf{P_o} + \mathbf{P_o}(A_o V_\theta)^T + A_o R A_o^T = 0.$$

II.3 SUMMARY OF THE RESULTS; A QUALITY CRITERION; CHOICE OF THE OPTIMUM GAIN MATRIX A.

Let us first summarize the results of the preceding paragraphs.

Theorem 3:

(i) Select A of the form

$$A = \gamma^{-\alpha} A_o \tag{2.30}$$

Then, when γ tends to zero, the asymptotically optimum α is given by

$\alpha = 0$ **in the zero drift case (2.17)**

$\alpha = 1/3$ **in the non zero drift case (2.26).** (2.31)

<u>(ii) zero drift case (2.17)</u>; **we distinguish two subcases:**

(ii-a) U_S **is asymptotically stable (the true system moves according to a second order process of finite energy); then, for** $t\gamma > \tau_o$ **, and** $\gamma \to 0$ **,we have**

$$E\|\theta_t - S_t\|^2 \simeq \gamma \; Tr(P_{22}) \tag{2.32}$$

where

$$\underline{P} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \tag{2.33}$$

is the solution of the Lyapunov equation:

$$\underline{F}\,\underline{P} + \underline{P}\,\underline{F}^T + \underline{Q} = 0$$

$$\underline{F} = \begin{bmatrix} U_S & 0 \\ -U_S & AV_\theta \end{bmatrix} , \quad \underline{Q} = \begin{bmatrix} Q & -Q \\ -Q & Q + ARA^T \end{bmatrix} \tag{2.34}$$

(ii-b) $U_S = 0$ **(the true system moves according to a random walk); here,**

$$E\|\theta_t - S_t\|^2 \simeq \gamma \; Tr(P) \tag{2.35}$$

where P **is the solution of the Lyapunov equation**

$$AV_\theta P + P(AV_\theta)^T + ARA^T + Q = 0. \tag{2.36}$$

<u>(iii) non zero drift case (2.26)</u>; **for** $t\gamma = \tau$

$$E\|\theta_t - S_t\|^2 \simeq \gamma^{2/3} \; Tr(\; (A_o V_\theta)^{-1} UU^T (A_o V_\theta)^{-T} + P_o \;), \quad U \equiv U(\bar{S}(\tau)), \tag{2.37}$$

where P_o **is the solution of the Lyapunov equation**

$$A_o V_\theta P_o + P_o (A_o V_\theta)^T + A_o R A_o^T = 0. \tag{2.38}$$

Comments: the main surprinsing point in the theorem is the point (i): it expresses the strong difference between the zero drift and the non zero drift case. In the former case, the matrix gain has to be of the same order of magnitude as the speed of variation of the true system, whereas a much larger gain matrix has to be chosen in the later case, when a bias has to be compensated for.

We shall use the results of the theorem 3 for introducing a quality criterion, which will solve the problem 2 of the section I.3.

Theorem 4: a quality criterion.

Assume that a model of the time variations of the true system has been given. Then, for $t\gamma = \tau > 0$ **and** $\gamma \to 0$**, we have**

$$E\|\theta_t - S_t\|^2 \approx \gamma^{1-\alpha} \quad \Phi(A_o V_\theta(\bar{S}),\ \Delta(\bar{S})),\ \bar{S} \equiv \bar{S}(\tau) \tag{2.39}$$

where

$$\Delta \equiv V_\theta^{-1}.R.V_\theta^{-T} \tag{2.40}$$

the exponent α **and matrix gain** A_o **being chosen according to the theorem 3; the function in (2.39) has the property that**

$$\Delta \to \Phi(A_1, \Delta) \tag{2.41}$$

is increasing when A_1 **is fixed. The matrix** Δ **thus entirely characterizes the tracking capability of the algorithm (i.e. of the random vector field** $V_t(\theta, S)$ **): we will refer to** Δ **as the quality criterion of the algorithm.**

PROOF: the result is an immediate consequence of the theorem 3, if we rewrite every formula of the theorem in terms of the normalized matrix gain

$$A_1 = A_o.V_\theta \tag{2.42}$$

Use of the quality criterion: the quality criterion can be used in the following way. Assume we want to compare two different random vector fields V^1 and V^2 for tracking the same system S . Then, if $\Delta^1 < \Delta^2$, we can claim that the algorithm 1 is more efficient for tracking than 2 in the following sense: given a model of the time variations of the true system, for every choice of the matrix gain in the algorithm 2, there exists a matrix gain for the algorithm 1 which produces a smaller variance to the tracking error (2.12).

A connection with the case of decreasing gain algorithms: we will here show that the same quality criterion plays also a role for decreasing gain algorithms. Recall that decreasing gain algorithms are used for identifying **constant** systems. typical algorithms are of the form

$$\theta_{t+1} = \theta_t + \frac{1}{t} A\ V_t(\theta_t)\ ,\quad \theta \in \mathbf{R}^n \tag{2.43}$$

Let us denote by θ_* the unique stable equilibrium of the ODE associated to this algorithm, and denote by V and R the quantities (1.18) evaluated at θ_*. We shall report here the results of Kushner & Huang (1979); a matrix gain A is said to be **admissible** if it satifies

$$(AV + \frac{I}{2}) \text{ is asymptotically stable.} \tag{2.44}$$

Then, under suitable conditions, we have the following central-limit theorem

$$t^{1/2}(\theta_t - \theta_*) \to \mathbf{N}(0,P_A) \tag{2.45}$$

where P_A is the solution of the Lyapunov equation

$$(AV_\theta + \frac{I}{2})\, P_A + P_A(AV_\theta + \frac{I}{2})^T + ARA^T = 0. \tag{2.46}$$

Then, it is easy to show that

$$\min \{P_A : \text{A admissible}\} = V_\theta^{-1} R\, V_\theta^{-T} \quad (= \Delta) \tag{2.47}$$

the minimum being realized with the gain

$$A_* = -V_\theta^{-1} . \tag{2.48}$$

Hence, the same criterion measures at the same time the convergence rate of the algorithm used on time invariant systems (with a decreasing gain), and the tracking capability of the algorithm used on time varying systems (with a constaint gain).

Now, the problem of the optimal choice of the matrix gain will be investigated. For this purpose, we must assume that **the user has some prior knowledge of a model of the time variations of the true system**; the purpose is then to design an optimum matrix gain according to this prior knowledge. It is clear that the theorem 3 gives an answer to this question: in each of the cases (ii) or (iii), **the problem of the optimum choice of the matrix gain can be reformulated as a minimization problem with a constraint given by a Lyapunov equation.** We were not able to formulate closed form solutions to these minimization problems in all the cases, but we shall give an answer for the case (ii-b), where the true system moves according to a random walk.

Theorem 5: optimum choice of the matrix gain.

Assume the true system moves according to a random walk (with possibly dependent increments), which corresponds to the case (ii-b) of theorem 3. Then, according to the notations of (ii-b), the optimum pair (A_*, P_*) is given by

$$\begin{aligned} &P_* . \Delta^{-1} . P_* = Q \\ &A_* = -\, P_* . V_\theta^T . R^{-1} = Q^{1/2} . R^{-1/2} \end{aligned} \tag{2.49}$$

where the last equality of (2.49) is valid for convenient choices of the square roots.

PROOF: set $A_1 = AV_\theta$; P is defined as a function of A_1 as the solution of the Lyapunov equation

$$AP + PA^T + A\Delta A^T + Q = 0 \tag{2.50}$$

and we thus have to minimize in (2.50) P with respect to A. For this purpose, set

$$A = A_* + \tilde{A} \ , \ P = P_* + \tilde{P} \ , \tag{2.51}$$

where A_* and P_* are defined in (2.49). Using (2.49) and (2.50), we get that $\tilde{P}$ is the solution of

$$A \tilde{P} + \tilde{P} A^T + \tilde{A}.\Delta.\tilde{A}^T = 0, \tag{2.52}$$

which ensures that $\tilde{P} \geq 0$ whenever the normalized matrix gain A is stable. This finishes the proof.

III APPLICATIONS.

III.1 BACK TO THE EXAMPLES.

The DPLL.

In Benveniste, Vandamme & Joindot (1979), the quality criterion for both loops defined in (1.5) has been calculated under the assumption of a small distorsion due to the channel, i.e.

$$\sum_{k \neq 0} |s_k| << s_0$$

The criterion is approximately the same for both loops in this case, and is given by

$$\Delta = \frac{1}{2} \sum_{k>0} \frac{|s_k - s^*_{-k}|^2}{s_0^2} + \frac{\sigma^2}{s_0^2} \tag{2.53}$$

where σ^2 denotes the power of the noise v. Hence, both loops are equivalent in the case of a small distorsion. As a matter of fact , this formula allowed the authors to evaluate the performance of these loops in the presence of a dispersive (i.e. non trivial) channel, a result which has never been obtained before.
Note that, in this case, the quality criterion is time invariant, even if there is a time varying phase shift due to the channel.

As a matter of fact, R and V_θ depend on the dispersive channel (s_k) and noise power only, and not on the time varying unknown phase shift to be estimated: **this allows the user to compute in advance the optimum gain according to the theorem 5.** This is a pleasant feature of the DPLL algorithm, but is by no means a general situation.

The Recursive Least Squares algorithm: a connection with Kalman Filtering.

Using the orthogonality conditions between the regression vector and the noise, we get

$$V_\theta(\bar{S}) = -\,\mathbf{E}(\phi\phi^T)_{|\bar{S}} \equiv -\,\Sigma(\bar{S})$$

$$R(\bar{S}) = \sum_{n\in\mathbf{Z}} \mathbf{E}(v_n\phi_n\phi_0^T v_0)_{|\bar{S}} = \sigma^2\,\Sigma(\bar{S}) \tag{2.54}$$

where σ^2 is the variance of the noise v: the terms with nonzero indices in the sum vanish since the noise is orthogonal to the regression vector. Note that, here, **these quantities depend on the trajectory $\bar{S}$ of the ODE of the true system,** an unpleasant feature which was not encountered in the DPLL. The quality criterion is then equal to

$$\Delta = \sigma^2\,\Sigma^{-1} \tag{2.55}$$

so that Δ is here nothing but **the Fisher information matrix.**

We shall now calculate the optimum matrix gain A according to the theorem 5. Note that, in this case, $\bar{S}$ is fixed, and so are R and V_θ. The optimal pair (A_*, P_*) is

$$P_*\,\Delta\,P_* = \sigma^2 Q \quad , \text{ or } \quad P_* = \sigma\, Q^{1/2}\,\Sigma^{-1/2}$$

$$A_* = \sigma^{-2}\,P_* = \sigma^{-1}\,Q^{1/2}\,\Sigma^{-1/2} \tag{2.56}$$

with convenient choices for the square roots. Note that the computation of the optimal gain requires the knowledge of Σ, which is precisely not known in advance, especially if the true system is time varying. We shall now show that **using an appropriate Kalman filtering formulation of the RLS algorithm provides us with an estimate of the optimal matrix gain (2.56).** This formulation is the following:

$$S_{t+1} = S_t + \gamma\, w_t \quad , \quad \mathrm{cov}\,(w) = Q$$

$$y_t = \phi_t^T S_t + v_t \tag{2.57}$$

Setting

$$\underline{Q} = \sigma^{-2}\,Q$$

the Kalman filter equations are

PART 3: MULTISTEP ADAPTIVE ALGORITHMS.

Although they have used a single name, the Soviet engineers have introduced in fact two different classes of multistep adaptive algorithms. The first class is a copy of the multistep methods of the numerical analysis; those methods will be shown to have no advantage over monostep algorithms, neither for the convergence rate (decreasing gain algorithms and constant systems) nor for the tracking capability, but they are reported to improve in some cases the transient behaviour of the algorithms (these claims are supported by simulations (Shil'man & Yastrebov (1976, 1978)). The second class is in fact wellknown in the community of digital communications for the case of the DPLL, where they are referred to as "higher order loops"; they will also be shown to be related to extended Kalman filtering with a prior model of the time variations of the true system, and it will be proved that they improve the tracking capability with respect to conventional monostep algorithms.

I HOW TO BUILD MULTISTEP ALGORITHMS .

I.1 MULTISTEP ALGORITHMS OF THE FIRST KIND.

They are extensively studied in Shil'man & Yastrebov (1976, 1978). Their goal is to force the algorithm (1.10) to be closer to the ODE (1.11), exactly as for the finite difference approximations of ordinary differential equations (Henrici (1963)). Using the z-transform notation, those algorithms are of the form

$$(I - z^{-1})\,\theta_t = \gamma\, A(z^{-1}).V_t(\theta_t) \qquad (3.1)$$

where $V_t(\theta)$ is as before, γ is the small parameter we shall let tend to zero, and

$A(z^{-1})$ is a stable rational transfer function (3.2)

The filters used by Shil'man & Yastrebov are all-pole, but rational filters can be used as well. It is easy to see that the ODE corresponding to (3.1) is nothing but

$$\dot{\bar\theta} = A(1).V(\bar\theta) \qquad (3.3)$$

On the other hand, defining

$$\begin{aligned} R(\bar\theta) &= \sum_{n\varepsilon Z} \mathrm{cov}(V_n(\bar\theta),V_0(\bar\theta)) \\ \underline{R}(\bar\theta) &= \sum_{n\varepsilon Z} \mathrm{cov}(\,A(z^{-1})V_n(\bar\theta),\ A(z^{-1})V_0(\bar\theta)\,), \end{aligned} \qquad (3.4)$$

we have the relationship

$$
\begin{aligned}
\underline{R}(\bar{\theta}) &= \sum_{n \varepsilon Z} \int e^{in\phi} A(e^{i\phi}) R(\bar{\theta}) A^T(e^{-i\phi}) \, d\phi \\
&= A(1)^{-1}.R(\bar{\theta}).A(1)^{-T}
\end{aligned}
\tag{3.5}
$$

Finally, (3.3) and (3.5) show that (3.1) and (1.10) are completely equivalent from the viewpoint of both the convergence rate and the tracking capability. This finishes the study of this class of multistep methods.

I.2 MULTISTEP ALGORITHMS OF THE SECOND KIND.

We refer the reader to Korostelev (1981) for very deep results on such methods from the viewpoint of the large deviation theory. The best way of obtaining these algorithms is to start with the ODE

$$\dot{\bar{\theta}} = V(\bar{\theta}) \tag{3.6}$$

Let $A(s)$ denote a stable continuous time rational transfer function defined by

$$A(s) = J + H(sI-F)^{-1}G \tag{3.7}$$

and replace (3.6) by

$$\dot{\bar{\theta}} = A(s).V(\bar{\theta}) \tag{3.8}$$

Write (3.8) in the state space form

$$\begin{pmatrix} \dot{\bar{\Theta}} \\ \\ \dot{\bar{\theta}} \end{pmatrix} = \begin{pmatrix} F & G \\ \\ H & J \end{pmatrix} \begin{pmatrix} \bar{\theta} \\ \\ V(\bar{\theta}) \end{pmatrix} \tag{3.9}$$

and form the adaptive algorithm corresponding to the extended ODE (3.9)

$$\begin{pmatrix} \Theta_{t+1} \\ \\ \theta_{t+1} \end{pmatrix} = \begin{pmatrix} \Theta_t \\ \\ \theta_t \end{pmatrix} + \gamma \begin{pmatrix} F & G \\ \\ H & J \end{pmatrix} \begin{pmatrix} \Theta_t \\ \\ V_t(\theta_t) \end{pmatrix} \tag{3.10}$$

which is also of the form

$$(I-z^{-1})\theta_t = \gamma\, A_\gamma(z^{-1}).V_t(\theta_t)$$

$$A_\gamma(z^{-1}) = J + \gamma\, H(\, zI - (I+\gamma F)\,)^{-1}G \qquad (3.11)$$

thus resulting in a discrete time filter depending upon γ. Again, the multistep methods analysed by Korostelev correspond to all-pole filters $A(s)$.

From now on, we shall concentrate on the multistep algorithms of the second kind, and we shall refer to these simply as "multistep algorithms".

I.3 ANALYSIS OF THE RELATED ODE.

This ODE is given by (3.8); the introduction of the filter $A(s)$ can modify the stability of the original ODE. An analysis of this problem is now given.

Theorem 6: stability of the multistep ODE.
Let us denote by θ_* a stable equilibrium point of the original ODE (3.6).
(i) Global stability: assume there exists a positive definite matrix Π such that

$$(\theta-\theta_*)^T.\Pi.(\theta-\theta_*) \qquad (3.12)$$

be a Lyapunov function of (3.6). Then a sufficient condition for (3.8) to be globally stable with θ_* as equilibrium is that

$\Pi.A(s)$ **be a strictly posisitve real transfer function** (3.13).

(ii) Local stability: a necessary and sufficient condition for θ_* to be a locally stable equilibrium of (3.8) is that the matrix

$$A(0).V_\theta(\theta_*) \qquad (3.14)$$

be asymptotically stable.

PROOF of (i): set $\underline{\theta} = \Pi\theta$. Then, (3.8) is rewritten as

$$\dot{\underline{\theta}} = \Pi\; A(s)\; V(\Pi^{-1}\underline{\theta}) \qquad (3.15)$$

Set $\tilde{\underline{\theta}} = \underline{\theta} - \underline{\theta}_*$; (3.12) implies

$$\tilde{\underline{\theta}}^T\; V(\Pi^{-1}\underline{\theta}) = \tilde{\theta}^T\; \Pi\; V(\theta) \;\leq\; 0, \text{ and } <0 \text{ if } \tilde{\underline{\theta}} \neq 0. \qquad (3.16)$$

Then, thanks to Popov's hyperstability criterion, (3.16) and (3.13) implie that the derivative of $\tilde{\underline{\theta}}$ converges to zero for every inital condition; but this implies also that $\dot{\theta}$ converges to zero, and also $V(\theta)$ by (3.6), which implies finally that θ converges to θ_*.

Proof of (ii): since F is asymptotically stable, (3.14) implies that the matrix

$$\begin{pmatrix} F & GV_{\theta} \\ H & JV_{\theta} \end{pmatrix} \tag{3.17}$$

is also asymptotically stable; but this is exactly the condition of local stability of the pair $(0, \theta_*)$ for the ODE (3.9). This finishes the proof of the theorem.

In the sequel, we shall assume that θ_* is at least a locally stable equilibrium of the ODE (3.8).

II TRACKING PROPERTIES OF MULTISTEP ALGORITHMS.

II.1 THE CASE OF THE DPLL.

A very common situation in the case of the DPLL is the simultaneous presence of jitter and frequency offset (Falconer (1976-a) and (1976-b)); a jitter corresponds to a motion of the phase shift p according to a sinusoïd with a known frequency, and a frequency offset corresponds to an approximately constant drift due to an incorrect tuning of the voltage controlled oscillator in the loop. A possible continuous time model for the evolution of the true phase shift p of the system is

$$\dot{p} = \frac{\sigma s}{s^2 + \varepsilon s + \omega^2} v_1 + \frac{1}{s} \sigma' v_2 + \mu \tag{3.18}$$

where ω is the frequency of the jitter, ε and σ are small parameters related to the amplitude of the jitter by

$$\sigma^2 / \varepsilon\omega^2 = \text{amplitude of the jitter} \tag{3.19}$$

and μ is the nominal value of the offset, and σ' is a small parameter reflecting the uncertainty on the value of the offset; finally, v_1 and v_2 are independent standard white noises. It is clear that (2.7) is a very poor model for reflecting this behaviour.

II.2 A MODEL FOR ANALYSING THE TRACKING PROBLEM.

In view of the preceding example, it is interesting to generalize the class of models (2.7), by assuming that the true system obeys the following dynamics

$$\begin{pmatrix} Z_{t+1} \\ S_{t+1} \end{pmatrix} = \begin{pmatrix} Z_t \\ S_t \end{pmatrix} + \begin{pmatrix} F_* & G_* \\ H_* & J_* \end{pmatrix} \begin{pmatrix} Z_t \\ U_t(S_t) \end{pmatrix} \tag{3.20}$$

where, for S fixed, $U_t(S)$ is as usually a stationary random vector field.

To this model, we shall associate the following multistep algorithm, where the dimensions of the matrices are the same respectively in (3.20) and (3.21)

$$\begin{pmatrix} \theta_{t+1} \\ \\ \theta_{t+1} \end{pmatrix} = \begin{pmatrix} \theta_t \\ \\ \theta_t \end{pmatrix} + \begin{pmatrix} F & G \\ \\ H & J \end{pmatrix} \begin{pmatrix} \theta_t \\ \\ V_t(\theta_t, S_t) \end{pmatrix} \qquad (3.21)$$

where, again, $V_t(\theta, S)$ denotes as usually the random vector field defining the algorithm, for which we shall also assume that **the assumtpions 1 are satisfied;** moreover, we shall assume that, **for every S fixed, the ODE**

$$\dot{\theta} = A(s)\ V(\bar{\theta}, \bar{S}) \qquad (3.22)$$

is asymptotically stable, with $\theta = S$ **as unique equilibrium.** This later condition defnies the class of **admissible filters** A(s) for the multistep algorithm.

Again, we are interested in the joint behaviour of the equations (3.21), (3.22) when the true system moves slowly, i.e. when $U_t(S)$ is small; thus, as for the study of monostep methods, we shall introduce a small parameter γ for the purpose of the asymptotic analysis. Setting

$$\underline{\underline{S}} = \begin{pmatrix} Z \\ \\ S \end{pmatrix},\ \underline{\underline{\theta}} = \begin{pmatrix} \theta \\ \\ \theta \end{pmatrix},\ \underline{\underline{U}}(\underline{\underline{S}}) = \begin{pmatrix} Z \\ \\ U(S) \end{pmatrix},\ V(\underline{\underline{\theta}}, \underline{\underline{S}}) = \begin{pmatrix} \theta \\ \\ V(\theta, S) \end{pmatrix},\ \underline{\underline{A}} = \begin{pmatrix} F & G \\ H & J \end{pmatrix} \qquad (3.23)$$

we can rewrite (3.21) and (3.22) with the small parameter γ as follows

$$\begin{aligned} \underline{\underline{\theta}}_{t+1} &= \underline{\underline{\theta}}_t + \gamma\ \underline{\underline{A}}\ \underline{\underline{V}}_t(\underline{\underline{\theta}}_t, \underline{\underline{S}}_t) \\ \underline{\underline{S}}_{t+1} &= \underline{\underline{S}}_t + \gamma\ \underline{\underline{A}}_*\ \underline{\underline{U}}_t(\underline{\underline{S}}_t) \end{aligned} \qquad (3.24)$$

which looks like the form (2.8) analysed in the second part. But, don't forget that we are interested in the behaviour of θ, and not of $\underline{\underline{\theta}}$!

II.3 MAIN THEOREMS.

We shall restrict ourselves to the subclass corresponding to (ii-b) in the theorem 3, for the following two reasons: first, the other cases are more involved, exactly as for monostep algorithms, and , second, this is the only case for which we were able to derive explicit formulas for the optimum filter A(s). We shall thus assume

$$\bar{S} = \text{constant}, \quad U_S(\bar{S}) = 0. \qquad (3.25)$$

Take care that the condition (3.25) do not imply that the pair (Z_t, S_t) moves according to a random walk, so that it is not possible to apply brute-force the results of the theorem 3 to the joint model (3.24)!

Since the mean values $(\bar{z},\bar{S})$ are constant (with $\bar{z} = 0$), we shall again delete the dependence of the forthcoming quantities with respect to these parameters. Because of the assumption (3.25), the contribution of the bias to (2.12) is equal to zero, so that the only problem is the evaluation of the variance. This evaluation will be obtained by applying the theorem 2 to the model (3.24). According to this theorem, the joint process $(\tilde{\underline{\underline{S}}},\tilde{\underline{\underline{\theta}}})$ is the solution of the following stochastic differential equation

$$\begin{pmatrix} d\underline{\underline{S}}_\tau \\ \\ d\underline{\underline{\theta}}_\tau \end{pmatrix} = \begin{pmatrix} F_* & 0 & 0 & 0 \\ H_* & 0 & 0 & 0 \\ 0 & G\,V_S & F & GV_\theta \\ 0 & J\,V_S & H & JV_\theta \end{pmatrix} \begin{pmatrix} \underline{\underline{S}}_\tau \\ \\ \underline{\underline{\theta}}_\tau \end{pmatrix} d\tau + \begin{pmatrix} \begin{pmatrix} G_* \\ J_* \end{pmatrix} Q^{1/2} & 0 \\ 0 & \begin{pmatrix} G \\ J \end{pmatrix} R^{1/2} \end{pmatrix} dW_\tau$$

Using (2.11), we get

$$\begin{pmatrix} d\underline{\underline{S}}_\tau \\ \\ d\underline{\underline{\theta}}_\tau \end{pmatrix} = \begin{pmatrix} F_* & 0 & 0 & 0 \\ H_* & 0 & 0 & 0 \\ 0 & -GV_\theta & F & GV_\theta \\ 0 & -JV_\theta & H & JV_\theta \end{pmatrix} \begin{pmatrix} \underline{\underline{S}}_\tau \\ \\ \underline{\underline{\theta}}_\tau \end{pmatrix} d\tau + \text{same as above} \qquad (3.27)$$

Finally, **the joint process** $(\tilde{\underline{\underline{S}}},\tilde{\underline{\underline{\theta}}}-\tilde{\underline{\underline{S}}})$ **is the solution of the following linear stochastic differential equation**

$$\begin{pmatrix} d\underline{\underline{S}}_\tau \\ \\ d\underline{\underline{\theta}}_\tau \end{pmatrix} = \begin{pmatrix} F_* & 0 & 0 & 0 \\ H_* & 0 & 0 & 0 \\ F-F_* & 0 & F & GV_\theta \\ H-H_* & 0 & H & JV_\theta \end{pmatrix} \begin{pmatrix} \underline{\underline{S}}_\tau \\ \\ \underline{\underline{\theta}}_\tau \end{pmatrix} d\tau + \begin{pmatrix} \begin{pmatrix} G_* \\ J_* \end{pmatrix} Q^{1/2} & 0 \\ -\begin{pmatrix} G_* \\ J_* \end{pmatrix} Q^{1/2} & \begin{pmatrix} G \\ J \end{pmatrix} R^{1/2} \end{pmatrix} dW_\tau \qquad (3.28)$$

We shall now assume that **the true system moves according to a process of finite energy,** i.e.

F_* is asymptotically stable, (3.29)

and we shall restrict ourselves to **admissibles filters** A(s) (cf.(3.22)). Thanks to these assumptions, the steady state covariance matrix of the pair $(\underline{\underline{S}}, \underline{\underline{\theta}}-\underline{\underline{S}})$, which we shall denote by $\underline{\underline{P}}$, is the solution of the Lyapunov equation

$$\underline{\underline{F}}\,\underline{\underline{P}} + \underline{\underline{P}}\,\underline{\underline{F}}^T + \underline{\underline{Q}} = 0 \tag{3.30}$$

where

$$\underline{\underline{F}} = \begin{pmatrix} F_* & 0 & 0 & 0 \\ H_* & 0 & 0 & 0 \\ F-F_* & 0 & F & GV_\theta \\ H-H_* & 0 & H & JV_\theta \end{pmatrix}$$

$$\underline{\underline{Q}} = \left(\begin{array}{c|c} \begin{pmatrix} G_* \\ J_* \end{pmatrix} Q\,(G_*^T, J_*^T) & -\begin{pmatrix} G_* \\ J_* \end{pmatrix} Q\,(G_*^T, J_*^T) \\ \hline -\begin{pmatrix} G_* \\ J_* \end{pmatrix} Q\,(G_*^T, J_*^T) & \begin{pmatrix} G_* \\ J_* \end{pmatrix} Q\,(G_*, J_*) + \begin{pmatrix} GV_\theta \\ JV_\theta \end{pmatrix} R\,((GV_\theta)^T, (JV_\theta)^T) \end{array}\right) \tag{3.31}$$

This formula allows us to extend the validity of the quality criterion we have defined before to the multistep algorithms:

Theorem 7: assume that a model of the time variations of the true system has been given according to (3.24), and that (3.25) is in force. Then, for $t\gamma=\tau>0$ **and** $\gamma \to 0$, **we have**

$$\mathrm{E}\,||\theta_t - S_t||^2 \simeq \gamma\ \Phi(\underline{\underline{A}}_1, \Delta) \tag{3.32}$$

where Δ **is given by (2.40), and**

$$\underline{\underline{A}}_1 = \begin{pmatrix} F & GV_\theta \\ H & JV_\theta \end{pmatrix} \tag{3.33}$$

The function Φ **in (3.32) has the property that**

$$\Delta \to \Phi(\underline{\underline{A}}_1, \Delta) \tag{3.34}$$

is increasing when $\underline{\underline{A}}_1$ **is fixed. This extends the validity of the quality criterion** Δ **to multistep algorithms.**

This theorem is an easy consequence of the formulas (3.30) and (3.31). This theorem and the proof of the theorem 4 allows us reasonably to claim that **our quality criterion is valid for general** $U_t(S)$ **as well**; a formal proof supporting this claim would be rather tedious, so that we did not include it in this paper.

We shall now investigate the problem of the optimal choice of the filter; as for the monostep case, we shall give an answer to this problem only when (3.25) is in force.

Theorem 8: <u>calculation of the optimal filter.</u>

Assume that the true system moves according to a given model (3.24) satisfying to (3.25). Define the filter $A_{opt}(s)$ as follows

$$A_{opt}(s) = J_{opt} + H_{opt}(sI - F_{opt})^{-1}G_{opt}$$
$$F_{opt} = F_* \,, \quad H_{opt} = H_* \tag{3.35}$$
$$G_{opt} = P_{12}\, V_\theta^T\, R^{-1} \,, \quad J_{opt} = P_{22}\, V_\theta^T\, R^{-1}$$

where

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \tag{3.36}$$

is the unique nonnegative definite solution of the following Algebraic Riccati Equation

$$P \begin{pmatrix} 0 & 0 \\ 0 & \Delta \end{pmatrix} P = \begin{pmatrix} F_* & 0 \\ H_* & 0 \end{pmatrix} P + P \begin{pmatrix} F_*^T & H_*^T \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} G_* \\ J_* \end{pmatrix} Q(G_*^T \; J_*^T) \tag{3.37}$$

Then, the filter $A_{opt}(s)$ **is admissible, and realizes the minimum of the criterion (2.12) among all the admissible filters.**

PROOF: see the appendix B.

III APPLICATIONS.

III.1 THE RLS ALGORITHM: A CONNECTION WITH KALMAN FILTERING.

We shall here assume that the true system moves according to the model

$$\begin{pmatrix} Z_{t+1} \\ S_{t+1} \end{pmatrix} = \begin{pmatrix} Z_t \\ S_t \end{pmatrix} + \gamma \begin{pmatrix} F_* & G_* \\ H_* & J_* \end{pmatrix} W_{t+1} \tag{3.38}$$

where W_t is a standard white noise. Add the regression equation

$$y_t = \phi_t^T S_t + v_t \tag{3.39}$$

and use the Kalman filter equations for computing the estimate Θ of the extended state defined in (3.38). This gives

$$\Theta_{t+1} = (\mathbf{F}_\gamma - K_t \Phi_t^T)\Theta_t + K_t y_t$$

$$K_t = \mathbf{F}_\gamma \Sigma_t \Phi_t (\Phi_t^T \Sigma_t \Phi_t + \sigma^2)^{-1}$$

$$\Sigma_{t+1} = \mathbf{F}_\gamma (\Sigma_t - \Sigma_t \Phi_t (\Phi_t^T \Sigma_t \Phi_t + \sigma^2)^{-1} \Phi_t^T \Sigma_t) \mathbf{F}_\gamma^T + \gamma^2 \mathbf{Q}$$

$$\mathbf{F}_\gamma = \begin{pmatrix} I + \gamma F_* & 0 \\ \gamma H_* & I \end{pmatrix} \qquad \Phi_t^T = (0 \ \phi_t^T) \tag{3.40}$$

$$\mathbf{Q} = \begin{pmatrix} G_* \\ J_* \end{pmatrix} (G_*^T \quad J_*^T)$$

Now, we shall proceed as in the analysis of (2.58): taking into account that, for γ small, the error covariance matrix is small and slowly time varying, we can 1/ neglect Σ with respect to I, 2/ use an averaging argument for

replacing $\Phi_t \Phi_t^T$ by its expectation $\mathbf{R} = \begin{pmatrix} 0 & 0 \\ 0 & R \end{pmatrix}$ (3.41)

Finally, we take a first order expansion of the so modified ARE, thus obtaining

$$\sigma^{-2} \Sigma \mathbf{R} \Sigma = \gamma (\mathbf{F}\Sigma + \Sigma \mathbf{F}^T) + \gamma^2 \mathbf{Q}$$

$$\mathbf{F}_\gamma = I + \gamma \mathbf{F} \ , \text{ i.e. } \ \mathbf{F} = \begin{pmatrix} F_* & 0 \\ H_* & 0 \end{pmatrix} \tag{3.42}$$

which is exactly the optimal filter given by (3.37).

Again, the conclusion is that **in order to compute the optimum filter for the multistep RLS algorithm, it is convenient to set the tracking problem as a Kalman filtering problem with the proper choice of the prior model of time variation of the true system.**

As for monostep algorithms, it is reasonable to expect that, when this is possible, **setting the tracking problem as an Extended Kalman Filtering problem with the proper choice of the prior model of the time variations of the true system will provide us asymptotically with the optimum filter for the multistep procedure.** Such an approach was for instance used by Gersh & Kitagawa (1984), where the following prior model is used for the time variations of the true system: $\nabla^k s$ = noise, where $\nabla s_t = s_{t+1} - s_t$, and the order k is selected according to an AIC criterion.

III.2 THE DPLL.

We shall investigate the case of the jitter, which corresponds to the continuous time model

$$\dot{p} = \frac{\sigma s}{s^2 + \varepsilon s + \omega^2} v \ , \quad v \text{ standard white noise,} \tag{3.43}$$

for the time variations of the true system, where ω is the frequency of the jitter, ε is a small positive parameter ensuring the stability of the system, and the standard deviation σ of the noise v_t is related to ε and ω through

$$\sigma^2/\varepsilon\omega^2 = \text{amplitude of the jitter} \tag{3.44}$$

The preceding theory applies with

$$F_* = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -\varepsilon \end{bmatrix}, \quad G_* = \begin{bmatrix} 0 \\ \sigma \end{bmatrix}, \quad H_* = (0 \quad 1), \quad J_* = 0$$

$$Q = \text{identity}, \quad \Delta \text{ given by (2.53)} \tag{3.45}$$

The optimum filter of the multistep algorithm (also known as the **loop filter**) is then computed using (3.45), (3.35,36,37), and (2.53); then, the algorithm is obtained through (3.8,9,10) where γ is nothing but the inverse of the sampling rate (i.e. the discretization step).

IV SOME GUIDELINES FOR THE DESIGN OF MULTISTEP ADAPTIVE ALGORITHMS.

Again, three problems are encountered.

Step 1: selection of the best algorithm, i.e. of the best random vector field $V_t(\theta,S)$: as we have shown in this third part, the quality criterion we have introduced is valid for evaluating the tracking capability of an algorithm for monostep and multistep procedures as well. Hence the same remarks and hints hold as before.

Step 2: choice of a prior model of the time variations of the true system. The best is to define this model in the continuous time domain. Two cases occur:

CASE 1: the prior knowledge on the true system is sufficient for allowing to choose reliably a higher order model of the form $S = A_*(s).U(S)$. This is for example the case for the DPLL, when the parameters of the jitter and offset are known. It should be acknowledged that the sensitivity of the performances of the algorithm to the choice of this prior model is rather unknown, and may be significant especially for multistep methods.

CASE 2: the prior knowledge on the time variations of the true system is very poor; in this case, the robustness of the method with respect to this uncertainty is a key issue. Then the most commonly used approach is a conservative one: select a random walk model and use a monostep procedure; Bohlin (1977) gives an interesting example of such an approach for the RLS algorithm where the covariance of the noise of the dynamics of the true system is also identified. On the other hand, a different attractive approach is used by Gersh & Kitagawa (1984) where a given family of models is used ($\nabla^k S_t$ = noise, ∇ = difference operator) with an unspecified order k; then this order is estimated using an AIC criterion.

In any case, the analysis we have developed clearely points out that the time dependency of the random vector field $U_t(S)$ defining the time variations of the true system has very little importance, since this dependency disappears in the diffusion model introduced in the theorem 2. This justifies the use of random walk models for describing time varying systems, even if the short time behaviour of a random walk is erratic from a theoretical viewpoint, while the motion of the true system is generally smooth.

Step 3: choice of the optimal filter A(s) for the algorithm. We have given some indications for the case of a general model of the time variations of the true system when a monostep procedure is used; and we have given precise results when the tracking error is reduced to a variance (the bias being zero), for both cases of monostep and multistep procedures. According to the theorem 8, when the tracking error reduces to a variance, proceed as follows:

* select a filter of the same structure as the model used for describing the time variations of the true system.

* apply the theorem 8 for computing the optimal filter $A_{opt}(s)$.

* Form the multistep adaptive algorithm according to (3.35,36,37), the small parameter being the sampling period.

In some case, the optimal filter can be computed off-line, thanks to a prior knowledge of the required quantities; such a situation is, for example, illustrated by the DPLL. But in other cases (like the RLS algorithm), the required quantities are not known in advance, but have to be estimated on-line; this difficullty is, for example, overcomed when the tracking problem is setted as an Extended Kalman filtering problem as it has been illustrated by the case of the RLS algorithm. Note that the theory of recursive stochastic algorithms extends the validity of the use of extended Kalman filtering, since discontinuous random vector fields can be used (with discontinuities arising for example from the differentiation of non smooth functions) in the observation equation; in fact, the effect of these discontinuities is smoothed out thanks to the averaging principle supporting the adaptive algorithms.

CONCLUSION.

The theoretical problems underlying the use of adaptive algorithms for tracking slowly time varying systems have been investigated. Multistep algorithms were introduced for improving the tracking capability, and a complete design methodology was given for selecting the optimal filter of a multistep algorithm when some prior information is available on the time variations of the true system. A quality criterion was introduced for evaluating a priori this tracking capability, regardless of the form of the time variations of the true system (provided that they are slow). This was possible since we used an asymptotic approach which considerably reduced the difficulty of the analysis. The counterpart of this advantage is that the domain of validity of the approximations we have used is not known (this drawback was pointed out in Macchi & Eweda (1984), and is in fact non neglectible if a rigorous approach is searched for; on the other hand, the results of these authors are of limited validity, and of poor help for the design problem).

APPENDIX A: PROOF OF THE THEOREM 3, ANALYSIS OF A TWO-TIME SCALE STOCHASTIC APPROXIMATION.

We have only to investigate the non-zero drift case. Rewrite (2.8) as follows

$$
\begin{aligned}
S_{t+1} &= S_t + \gamma\, U_t(S_t) \\
\theta_{t+1} &= \theta_t + \gamma/\varepsilon \;\; A_\circ V_t(\theta_t, S_t)
\end{aligned}
\tag{A.1}
$$

$\gamma \to 0, \quad \varepsilon \to 0, \quad \eta = \gamma/\varepsilon \to 0$

We shall only be able to present a heuristic analysis of (A.1). The key idea in the analysis of (A.1) is the following. **The deviation $\theta_t - S_t$ is asymptotically the sum of two terms: the contyribution of the bias and the contribution of the variance; in order to minimize this deviation, one has to balance between these two contributions;** this will determine the relative order of magnitude of the two gains ε and γ.

First order approximant: the ODE.

As in the theorem 1, we shall first approximate in the slow time scale

$$\tau = \gamma t \tag{A.2}$$

the pair $(S_t, \theta_t)_{0<t<S/\gamma}$ where $S < \infty$ is fixed, by the coupled two-time scale ODE

$$
\begin{aligned}
\dot{\bar{S}} &= U(\bar{S}) \\
\varepsilon\, \dot{\bar{\theta}} &= A_\circ\; V(\bar{\theta}, \bar{S})
\end{aligned}
\tag{A.3}
$$

and we shall apply to (A.3) a singular perturbation analysis argument (Hoppenstaedt (1971)). Thanks to the assumption 2, for γ tending to zero, (A.3) is expanded into the following coupled static/dynamic system

$$
\begin{aligned}
\dot{\bar{S}} &= U(\bar{S}) \\
\bar{\theta} &\simeq \bar{S} + \varepsilon\, (A_\circ V_\theta(\bar{S}, \bar{S}))^{-1}.U(\bar{S}, \bar{S})
\end{aligned}
\tag{A.4}
$$

for $0 < \tau < S$.

Second order approximant: the diffusion approximation.

We shall proceed as in the theorem 2, but with convenient modifications. We shall consider the deviation

$$(\theta_t - \bar{\theta}(\tau),\; S_t - \bar{S}(\tau))$$

in the **slow** time scale (A.2), but with the normalization associated with the **fast** dynamics ηt. Set

$$(\tilde{S}^{\gamma}_{\tau}\ ,\ \tilde{\theta}^{\gamma}_{\tau}\ ) \equiv (\eta)^{-1/2}\ (S_t-\bar{S}(\tau)\ ,\ \theta_t-\bar{\theta}(\tau)\ ) \tag{A.5}$$

Then, assuming $\varepsilon \to 0$, $\tilde{S}^{\gamma}_{\tau}$ tends to zero in probability, whereas the fast diffusion $\tilde{\theta}^{\gamma}_{\tau}$ lies in its steady state behaviour, given by

$$\tilde{\theta}_{\tau}\ \sim \mathbf{N}(0,P(\bar{S}(\tau))) \tag{A.6}$$

where P(S) is the solution of the Lyapunov equation

$$A_{\circ}V_{\theta}(S).P(S) + P(S).(A_{\circ}V_{\theta}(S))^{T} + A_{\circ}.R(S).A_{\circ}^{T} = 0 \tag{A.7}$$

Combining the two approximants.

Finally, for t large and γ small, (A.4-7) give together

$$\begin{aligned} \theta_t - S_t \simeq\ & \bar{\theta}(\tau) - \bar{S}(\tau) \\ & + \eta^{1/2}\ (\tilde{\theta}(\tau) - \tilde{S}(\tau)) \\ \simeq\ & \bar{\theta}(\tau) - \bar{S}(\tau) + \eta^{1/2}\ \tilde{\theta}(\tau) \\ \simeq\ & \varepsilon\ (A_{\circ}V_{\theta})^{-1}.U + \eta^{1/2}\ \tilde{\theta} \end{aligned} \tag{A.8}$$

which results in the optimum choice

$$\varepsilon.(\eta)^{-1/2} \to 1 \quad \text{when} \quad \gamma \to 0 \tag{A.9}$$

which gives exactly (2.31), whereas (A.6-8) give together (2.37,38).

This finishes the heuristic justification of the theorem 3. A rigorous analysis of the two-time scales stochastic approximation of the form (A.1) would be of great interest (for some data transmission systems, for example), but is far beyond the scope of this paper.

APPENDIX B: CALCULATION OF THE OPTIMAL FILTER OF THE MULTISTEP METHOD, PROOF OF THE THEOREM 8.

Step 1: convexity of the solution of the Lyapunov equation (3.30,31).

Rewrite (3.30,31) as follows

$$\underline{AP} + \underline{PA}^T + \underline{BP} + \underline{PB}^T + \underline{ARA}^T + \underline{Q} = 0 \tag{B.1}$$

where

$$\underline{B} = \begin{bmatrix} F_* & 0 & 0 & 0 \\ H_* & 0 & 0 & 0 \\ -F_* & 0 & 0 & 0 \\ -H_* & 0 & 0 & 0 \end{bmatrix}, \quad \underline{A} = \underline{\underline{F}} - \underline{B}, \quad \underline{P} = \underline{\underline{P}}$$

$$\underline{Q} = \left[\begin{array}{c|c} \begin{pmatrix} G_* \\ J_* \end{pmatrix} Q\ (G_*^T, J_*^T) & -\ \text{same} \\ \hline -\ \text{same} & +\ \text{same} \end{array}\right], \quad \underline{R} = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \begin{pmatrix} GV_\theta \\ JV_\theta \end{pmatrix} R\ ((GV_\theta)^T, (JV_\theta)^T) \end{array}\right] \tag{B.2}$$

We shall prove that

the function $\underline{A} \to \underline{P}$ is convex (B.3)

Since $\underline{B}$ satisfies $\underline{BR}=0$, by considering $\underline{A}+\underline{B}$ instead of $\underline{A}$, it is equivalent to prove that

the function $\underline{A} \to \underline{P}$ is convex, where

$$\underline{AP} + \underline{PA}^T + \underline{ARA}^T + \underline{Q} = 0 \tag{B.4}$$

For this purpose, set

$\underline{P} = \underline{P}' + \underline{P}''$, where

$$\underline{AP}' + \underline{P}'\underline{A}^T + \underline{Q} = 0$$
$$\underline{AP}'' + \underline{P}''\underline{A}^T + \underline{ARA}^T = 0 \tag{B.5}$$

It is sufficient to prove that

$\underline{A} \to \underline{P}'$ and $\underline{A} \to \underline{P}''$ are convex (B.6)

But, on one hand, the former is known to be convex, and on the other hand, the later is the composition of the two following convex maps

$\underline{A} \rightarrow \underline{A}^{+}$ (pseudo-inverse of $\underline{A}$)

and

$\underline{A}^{+} \rightarrow \underline{P}''$, where $(\underline{A}^{+})^{T}\underline{P}'' + \underline{P}''\underline{A}^{+} + \underline{Q} = 0.$

This finishes the proof of the step 1. The next steps will be devoted to prove that the optimum filter of the theorem 8 corresponds to a stationary point of the function (B.1).

Step 2: Among the filters of the form (H_*,F_*,G,J), the optimum filter is A_{opt} defined in the theorem 8.

For $F=F_*$, $H=H_*$, the equation (3.28) reduces to the following model for the process $(\tilde{\underline{\theta}}-\tilde{\underline{\underline{S}}})$:

$$\begin{bmatrix} d\theta_\tau \\ \\ d\theta_\tau \end{bmatrix} = \begin{bmatrix} F_* & GV_\theta \\ \\ H_* & JV_\theta \end{bmatrix} \begin{bmatrix} \theta_\tau \\ \\ \theta_\tau \end{bmatrix} d\tau + \begin{bmatrix} G \\ \\ J \end{bmatrix} R^{1/2}\, dW_\tau - \begin{bmatrix} G_* \\ \\ J_* \end{bmatrix} Q^{1/2}\, dV_\tau \tag{B.7}$$

Then the step 2 is a trivial consequence of the following lemma about Lyapunov equations

Lemma B.1: consider the following Lyapunov equation

$$AP + PA^{T} + A\underline{R}A^{T} + BP + PB^{T} + \underline{Q} = 0 \tag{B.8}$$

where the matrix A is constrained as follows

$$A+B \text{ is stable} \tag{B.9}$$

$$A = A'\underline{R}^{+} \text{for some } A' \tag{B.10}$$

where $\underline{R}^{+}$ is the pseudo inverse of $\underline{R}$. Then,

$P_* = \min_A P$

is the solution of the Riccati equation

$$P_*\underline{R}\ P_* = BP_* + P_*B^{T} + \underline{Q} \tag{B.11}$$

whereas the corresponding optimum gain matrix A_* is

$$A_* = -P_*\underline{R}^{+} \tag{B.12}$$

PROOF: easy verification left to the reader.

Apply this lemma to

$$B = \begin{bmatrix} F_* & 0 \\ H_* & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & GV_\theta \\ 0 & JV_\theta \end{bmatrix} \tag{B.13}$$

P is the covariance of the process $(\tilde{\underline{\theta}} - \tilde{\underline{\underline{S}}})$, and $\underline{Q},\underline{R}$ are defined in an obvious way to get the step 2.

A consequence of the step 1 is that the map

$$A(s) \rightarrow \mathbf{P} = \mathrm{cov}(\underline{\underline{\tilde{\theta}}} - \underline{\underline{\tilde{S}}})$$

is convex. Thanks to the step 2, it remains only to prove that the restriction of this map to $G = G_{opt}$, $J = J_{opt}$ admits a stationary point at the pair $(H,F) = (H_*,F_*)$. This will be done in the two following steps.

Step 3: for A_{opt} given as in the theorem 8, then

$$\mathrm{cov}(\underline{\underline{\tilde{\theta}}}\ ,\ \underline{\underline{\tilde{\theta}}} - \underline{\underline{\tilde{S}}}) = 0 \tag{B.14}$$

PROOF: the joint process $(\underline{\underline{\tilde{\theta}}}\ ,\ \underline{\underline{\tilde{\theta}}} - \underline{\underline{\tilde{S}}})$ satisfies the following coupled linear stochastic differential equations:

$$\begin{pmatrix} d\Theta_\tau \\ d\theta_\tau \end{pmatrix} = \begin{pmatrix} F & 0 \\ H & 0 \end{pmatrix}\begin{pmatrix} \Theta_\tau \\ \theta_\tau \end{pmatrix} d\tau + \begin{pmatrix} 0 & GV_\theta \\ 0 & JV_\theta \end{pmatrix}\begin{pmatrix} \Theta_\tau - Z_\tau \\ \theta_\tau - S_\tau \end{pmatrix} d\tau + \begin{pmatrix} G \\ J \end{pmatrix} R^{1/2}\, dW_\tau$$

$$\begin{pmatrix} d(\Theta-Z)_\tau \\ d(\theta-S)_\tau \end{pmatrix} = \begin{pmatrix} F_* & GV_\theta \\ H_* & JV_\theta \end{pmatrix}\begin{pmatrix} (\Theta-Z)_\tau \\ (\theta-S)_\tau \end{pmatrix} d\tau + \begin{pmatrix} F-F_* & 0 \\ H-H_* & 0 \end{pmatrix}\begin{pmatrix} \Theta_\tau \\ \theta_\tau \end{pmatrix} d\tau + \begin{pmatrix} G \\ J \end{pmatrix} R^{1/2} dW_\tau - \begin{pmatrix} G_* \\ J_* \end{pmatrix} Q^{1/2} dV_\tau \tag{B.15}$$

Then, the steady state covariance matrix Π of this process is the solution of the following Lyapunov equation

$$\underline{A}\,\Pi + \underline{\Pi}\,\underline{A}^T = \underline{Q}' \tag{B.16}$$

where

$$\underline{A} = \begin{pmatrix} F & 0 & F-F_* & GV_\theta \\ H & 0 & H-H_* & JV_\theta \\ 0 & 0 & F_* & GV_\theta \\ 0 & 0 & H_* & JV_\theta \end{pmatrix}, \qquad \begin{bmatrix} \begin{pmatrix} G \\ J \end{pmatrix} R(G^T,J^T) & \text{same} \\ \text{same} & \text{same} + \begin{pmatrix} G_* \\ J_* \end{pmatrix} Q(G_*^T,J_*^T) \end{bmatrix} \tag{B.17}$$

Setting $F=F_*$, $H=H_*$ in (B.17) and using the last formulas of (3.35) gives finally (B.14).

Step 4: The restriction of the map

$$A(s) \rightarrow \mathbf{P} = \mathrm{cov}(\underline{\underline{\tilde{\theta}}} - \underline{\underline{\tilde{S}}})$$

to $G=G_{opt}$, $J=J_{opt}$ **admits a stationary point at the pair** $(H,F) = (H_*,F_*)$.

PROOF: denote by $\underline{A}_*$ and Π_* respectively the pair $(\underline{A},\Pi)$ of (B.16) when $A_{opt}(s)$ is used. The derivative at $\underline{A}_*$ of the map $\underline{A} \rightarrow \Pi$ is the linear map $\delta\underline{A} \rightarrow \delta\Pi$ defined by

$$\underline{A}_*.\delta\underline{\Pi} + \delta\underline{\Pi}.\underline{A}_*^T + \delta\,\underline{A}.\underline{\Pi}_* + \underline{\Pi}_*.\delta\,\underline{A}^T = 0 \tag{B.18}$$

But, using (B.14) and (B.17), we get that

$$\delta\underline{\mathbf{A}}.\,\Pi_* \quad \text{is of the form} \quad \begin{pmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & 0 & 0 \\ \bullet & \bullet & 0 & 0 \end{pmatrix} \tag{B.19}$$

But this gives that the covariance **P** of the error $\tilde{\underline{\underline{\theta}}}-\tilde{\underline{\underline{S}}}$ satisfies

$$\begin{pmatrix} F_* & G_{opt}V_\theta \\ H_* & J_{opt}V_\theta \end{pmatrix} \delta\mathbf{P} + \delta\mathbf{P} \begin{pmatrix} F_* & G_{opt}V_\theta \\ H_* & J_{opt}V_\theta \end{pmatrix}^T = 0 \tag{B.20}$$

which implies $\delta\mathbf{P} = 0$, since the extended optimum matrix gain is asymptotically stable. This finishes the proof of this step, and, by the way, the proof of the theorem.

REFERENCES

A.BENVENISTE, P.VANDAMME, M.JOINDOT (1979): "Evaluation of the tracking capability of a digital phaselock loop in the presence of dispersive channels", Rep. NT/MER/TSF/1/CNET

A.BENVENISTE, M.GOURSAT, G.RUGET (1980): "Analysis of stocahstic approximation schemes with dependent and discontinuous forcing terms" IEEE-AC 25 n°5

A.BENVENISTE, G.RUGET (1982): "A measure of the tracking capability of recursive stochastic algorithms with constant gains", IEEE-AC 27 n°3

T.BOHLIN (1976): "Four cases of identification of changing systems", in System identification, advances and cases studies, R.K.Mehra & D.Lainiotis Ed., Academic Press, N-Y

D.P.DEREVITSKII & A.L.FRADKOV (1974): "Two models for analysing the dynamics of adaptation algorithms", Automation & Remote control, vol 35 n°1, 59-67

D.D.FALCONER (1976-a): "Jointly adaptive equalization and carrier recovery in two-dimensional digital communication systems" Bell Syst Tech J. voll 55, Mar 1976

D.D.FALCONER (1976-b): "Analysis of a gradient algorithm for simultaneous passband equalization and carrier phase recovery" Bell Syst Tech J. vol 55, Apr 1976

W.GERSH , KITAGAWA (1984): "A smoothness priors-time varying AR coefficient modeling of nonstationary covariance time series", 1984-CDC, Las Vegas, Dec 1984.

P.HENRICI (1963): "Error propagation for difference methods" Wiley & Sons, N-Y, London

F.HOPPENSTAEDT (1971): "Properties of solutions of ordinary differential equations with small parameters" Comm. on pure and Appl. Math. vol XXIV 807-840

A.P.KOROSTELEV (1981): "Multistep procedures of stochastic optimization" Avtomatikha i Telemekhanika n°5, 82-90, May 1981

H.J.KUSHNER, H.HUANG (1981): "Asymptotic properties of stochastic approximations with constant coefficients" SIAM J. Contr. & Optimiz.vol 19 n°1, 87-105

H.J.KUSHNER, H.HUANG (1982): "Diffusion approximation for the analysis of digital phaselocked loops" IEEE-IT 28 n°2

L.LJUNG, T.SODERSTROM (1983): "Theory and practice of recursive identification", MIT Press

O.MACCHI, E.EWEDA (1981): "Poursuite adaptative du filtrage optimal non stationnaire" CRAS, Série 1, t.293, 497-500

O.MACCHI, E.EWEDA (1984): "Tracking properties of adaptive nonstationary filtering" to appear IEEE-IT

M.METIVIER, P.PRIOURET (1984): "Application of a lemma of Kushner and Clark to general classes of stochastic algorithms" IEEE-IT, Mar 1984, special issue on adaptive filtering

A.RUSZCZYNSKI, W.SYSKI (1983): "Stochastic approximation method with gradient averaging for unconstrained problems" IEEE-AC 28 n°12

S.V.SHIL'MAN, A.I.YASTREBOV (1976): "Convergence of a class of multistep stochastic adaptation algorithms" Avtomatikha i Telemekhanika n°8 111-118, August 1976

S.V.SHIL'MAN, A.I.YASTREBOV (1978): "Properties of a class of multistep gradient and pseudogradient algorithms of adaptation and learning" Avtomatikha i Telemekhanika n°4 95-104, April 1978

B.WIDROW, J.M.McCOOL, M.G.LARIMORE, C.R.JOHNSON (1976): "Stationary and nonstationary learning characteristics of the LMS adaptive filter" Proc of the IEEE 64 n°8

B.WIDROW, E.WALACH (1984): "On the statistical efficiency of the LMS algorithm with nonstationary inputs" IEEE-IT, Mar.1984, Special issue on adaptive filtering